

In Brief: Statistics in Brief

Confidence Intervals

What is the Real Result in the Target Population?

Frederick J. Dorey PhD

Published online: 8 June 2010
© The Association of Bone and Joint Surgeons® 2010

Background

Authors of a recent article [1] indicated there was a difference ($p = 0.015$) in compliance at 6 months between the percentage of patients with osteoarthritis receiving insoles (71%; 32 of 45) versus the patients receiving braces (45%; 21 of 46). The small p value along with randomization of the patients provided strong evidence that the percent compliance in the target population (ie, all patients with osteoarthritis being treated by these methods) really is different. However, the p value alone does not provide us with much information concerning the magnitude of the difference in compliance in the target population. For that, we need to consider the estimated difference of 26% (71%–45%) and appropriate confidence interval.

Question

What is a confidence interval and how should it be used?

Discussion

The confidence interval provides us with a range of values that we believe, with a given level of confidence, contains

Each author certifies that he or she has no commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article.

F. J. Dorey (✉)
Department of Pediatrics, Children's Hospital Los Angeles,
4650 Sunset Blvd, Mailstop 54, Los Angeles, CA 90027, USA
e-mail: fdorey@chla.usc.edu

the true value of the statistical parameter (difference of percentages in this case) in the target population. The estimated parameter value of 26% difference in this small sample is the best point estimate we can make. However, we can be fairly certain this point estimate is not equal to the targeted true value as a very small sample rarely would give the exact estimate of the true proportion in the target population. Thus an interval estimate is needed to show the reliability of this point estimate. This is given by a confidence interval. To form a confidence interval, we first must decide on a level of confidence. The usual level of confidence taken is 95% but there is no requirement that it must be 95%. In this case, using a normal approximation method, the 95% confidence interval for the true difference is between 5.4% and 45.5%. Thus, the best we can say about the percent compliance in the target population, with 95% confidence, is the true difference in the compliance rate between the two treatments lies somewhere in the interval from 5.4% to 45.5%. Thus our conclusion is that the difference might be very large (as much as 45%) or very small (only approximately 5% or so). Despite the very small p value of 0.015, the real information coming from this study might not be as definitive as initially thought. Every parameter estimated in a study should have an estimated standard error. Although the standard deviation indicates how much variability there is in the sample data, the standard error gives us a measure of how precise the point estimate is to the true value in the target population. For example if we use the sample mean as an estimate for the true mean in the target population, the standard error is simply the standard deviation of the sample divided by the square root of the sample size. A 95% confidence interval is approximately plus or minus two standard errors from the parameter estimate. Thus, when reading the literature, taking plus or minus two standard errors is a quick way to

estimate a 95% confidence interval when only the standard error is provided.

The sample size is one of the most important issues related to the width of a confidence interval, and how much information is contained in the results from an experiment. For example, if the sample size had been 100 in each group in this experiment, the 95% confidence interval would have been from 12.8% to 39.2%. So a larger sample size results in a smaller width of the confidence interval and consequently in a more precise estimate of the true value in the target population.

One way to consider the confidence interval is the following: if we were to construct a large number of 95% confidence intervals from a large number of replications of a given experiment, 95% of them would in fact contain the true value. This is the basis for our 95% confidence that this particular interval does contain the true value. Choosing a different confidence level usually does not change the interpretation of the results very much. For example, the 90% confidence interval in the example above is 9.5% to 42.4%, which is not really all that different from the 95% confidence interval of 5.5% to 45.5%. Both intervals indicate that the true difference in compliance might be very large or very small. Thus, for example, if we were interested in knowing if there might be a large difference in compliance, both confidence intervals suggest that might be the case. With a 90% versus 95% confidence interval we gain information because we end up with a confidence interval of smaller width, but we lose some confidence that the interval contains the true value.

Myths and Misconceptions

A confidence interval is not a probability, and therefore it is not technically correct to say the probability is 95% that a given 95% confidence interval will contain the true value of the parameter being estimated. Since the event

(construction of the confidence interval) has already occurred, the true value of the targeted parameter either does or does not lie in the interval. A second misconception is the belief that if two confidence intervals are constructed separately for two groups and they overlap slightly, it must mean that the parameter values being estimated cannot be statistically significant based on a resulting test. However, as statistical tests usually are calculated based on the difference in the point estimates under certain assumptions, it is possible that the test could produce a statistically significant result even though the separate intervals overlapped. But in this case we still would be able to say that the true difference might not be very large owing to the overlapping confidence intervals.

Conclusions

The confidence interval gives us an interval that we believe, with a given level of confidence, contains the true value of the parameter being estimated in the target population. When evaluating several papers discussing a common subject, comparing confidence intervals may be a better way of summarizing the results than simply comparing the estimates and the p values. The confidence interval takes into account possible differences in the sample sizes of the papers involved. If there is considerable overlap in the confidence intervals from different studies, it becomes clear, despite possible differences in the point estimates, the results may not really differ very much.

Reference

1. van Raaij TM, Reijman M, Brouwer RW, Bierma-Zeinstra SM, Verhaar JA. Medial knee osteoarthritis treated by insoles or braces: a randomized trial. *Clin Orthop Relat Res.* 2010 February 23 [Epub ahead of print].